



TITLE:

# Machine Learning Approaches for Personalized Clinical Risk Modeling( Abstract\_要旨 )

AUTHOR(S):

Nori, Nozomi

---

CITATION:

Nori, Nozomi. Machine Learning Approaches for Personalized Clinical Risk Modeling. 京都大学, 2017, 博士(情報学)

ISSUE DATE:

2017-03-23

URL:

<https://doi.org/10.14989/doctor.k20504>

RIGHT:

許諾条件により本文は2018-03-20に公開; 許諾条件により要旨は2018-03-20に公開; 3章は同著者らによる論文1.  
Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD). 2015.と  
2. 電子情報通信学会論文誌和文D, Vol.J100-D, No.2, 2017に基づく。Copyright (C) 2017 IEICE. 4章は同著者らによる論文  
3. Proceedings of the 31st National Conference on Artificial Intelligence (AAAI). 2017.に基づく。

( 続紙 1 )

京都大学	博士（情報学）	氏名	則 のぞみ
論文題目	Machine Learning Approaches for Personalized Clinical Risk Modeling （機械学習による個別化臨床リスクモデリング）		
<p>（論文内容の要旨）</p> <p>本論文は、医療分野における患者の臨床リスク予測に対する機械学習の適用、特に疾病の多様性に起因する個別化リスクモデリングの困難さに対処するための方法について研究した結果をまとめたものであり、6章から構成されている。</p> <p>第1章は序論であり、本研究の目的とその内容を概観している。医療ビッグデータの出現とその利活用の必要性、臨床リスク予測の重要性について述べるとともに、本研究のテーマである疾病ごとに個別化された予測モデリングの必要性と、データ不足に起因する、精度のよいモデル獲得の困難さという課題を提示している。</p> <p>第2章では、関連研究についてまとめている。本研究が主に対象としているICU（集中治療室）のデータを対象とした生命リスクのモデリングについての研究事例、本研究の技術的背景となる機械学習のアプローチであるマルチタスク学習と転移学習についての既存研究、またこれらの医療データを対象とした取り組みの3つの観点から関連研究を整理している。</p> <p>第3章では、個々の疾病ごとに個別化された予測モデルの獲得を目指し、マルチタスク学習の考え方に基づく、医療ドメインの知識を有効利用する学習手法を提案している。本来、疾病や治療といった文脈によって予測モデルは異なるはずであるが、殆どの疾病では患者数が少ないため予測精度の高いモデルの獲得に必要なデータ数が確保できない。そこで、本研究では複数のモデルを同時に学習するマルチタスク学習の枠組みにおいて、ICD体系に基づく異なる疾病間の階層的な関係や、医薬品コードなどから導かれる異なる治療の間の類似性といった医療ドメイン知識を、学習の事前知識として正則化項に取り込むことによって、この問題の解決を図っている。実際にICUの実データを用いた比較実験によって、提案手法が既存のマルチタスク学習法を含む他手法と比較しても良好な予測精度を実現することを確認している。</p> <p>第4章では、疾病ごとに個別化された予測モデリングの考え方を、疾病の組み合わせにまで拡張した、より精密な個別化モデル獲得の方法を提案している。第3章で提案した方法を疾病の組み合わせ対して直接適用すると、個々の組み合わせに対応する患者数が極めて少数になってしまうという問題が生じる。そこで、本研究では疾病ごとのモデルパラメータを並べた行列に低ランク性を仮定することによって、疾病の組み合わせに対応する潜在的な表現を獲得するような新たなマルチタスク学習手法を提案している。第3章と同様の実データを用いた実験では、予測精度に一層の改善がみられることが示されている。</p> <p>第5章では、特定の疾病、とりわけ患者数の少ない低頻度疾病に対する予測モデルの予測精度を向上するための方法を提案している。第4章までに検討したマルチタスク学習は、全ての疾病あるいはその組み合わせに対して予測精度を上げることを目的としているため、特に低頻度疾病に対する予測精度を向上したいといった目的には必ずしも適さない。本研究ではこの問題により適した転移学習の枠組みによって解決</p>			

を行うことを提案している。ICUの実データを用いた比較実験では、提案アプローチが低頻度疾患に対する予測の再現率を特に向上することが示されている。

第6章は結論であり、本論文で得られた成果を要約している。即ち本論文は、医療分野における個別化リスクモデリングの方法について研究したものであり、マルチタスク学習の枠組みに基づき、医療ドメインの知識を適切に取り込む、あるいは疾病の組み合わせを考慮することでモデルの予測精度を高め、また、転移学習によって特定の低頻度疾患に対する予測精度の向上を実現したものである。本研究の成果は、ICUをはじめとする医療の現場においてデータに基づく意思決定を支援し、また、多様な疾病や患者に対する理解を一層進める可能性があるとして本論文を結んでいる。

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し  
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

本論文は、医療分野における患者の臨床リスク予測に対する機械学習の適用、特に疾病の多様性に起因する個別化リスクモデリングの困難さに対処するための方法を提案したものであり、得られた主な成果は次の通りである。

1. 疾病ごとに個別化された予測モデルの獲得に際し、機械学習におけるマルチタスク学習の考え方に基づく、医療ドメインの知識を有効利用する手法を提案している。ICD体系に基づく異なる疾病間の階層的な関係や、医薬品コードなどから導かれる異なる治療の間の類似性といった医療ドメイン知識を、マルチタスク学習の事前知識として学習の正則化項に取り込むことによって、個別化モデリングにおけるデータ過疎の問題を解決している。ICU（集中治療室）の実データを用いた比較実験によって、提案手法が他手法と比較して良好な予測精度を実現することが示されている。

2. 前述の疾病ごとに個別化された予測モデリングをさらに推し進め、疾病の組み合わせを考慮する方法を提案している。疾病の組み合わせを考えることによって、より精密なモデル化が期待できる反面、個々の組み合わせに対するデータの量は一層少なくなるという問題に対して、疾病の組み合わせに対応する潜在的な表現を行列分解の形でモデルパラメータに組み込むことで解決するマルチタスク学習手法を提案している。前述と同様の実データを用いた実験では、予測精度に一層の改善がみられることが示されている。

3. 低頻度疾病に対する予測精度を向上するための方法を提案している。前述の方法は、すべての疾病あるいはその組み合わせに対して予測精度を上げることを目的としているのに対し、特に低頻度疾病に対する予測モデルの精度を上げることを目的としている。この問題に対して、機械学習の転移学習の枠組みによって解決を行うことを提案している。同じくICUの実データを用いた比較実験では、提案アプローチが低頻度疾患に対する予測の再現率を特に向上することが示されている。

以上、本論文は医療分野における患者のリスクモデル化において、個別化されたモデルを得るための機械学手法について研究した結果をまとめたものであり、学術上、實際上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、平成29年2月28日に実施した論文内容とそれに関連した口頭試問の結果、合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。  
更に、試問の結果の要旨（例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。  
要旨公開可能日：2018年 3月 20日以降